DOCUMENT RESUME

ED 066 788 EA 004 380

TITLE	Evaluation of Instructional Programs. A Study.
INSTITUTION	Human Resources Research Council, Edmonton (Alberta).
SPONS AGENCY	Alberta Commission on Educational Planning, Edmonton.
PUB DATE	May 71
NOTE	60p.; Prepared for the Educational Planning Mission, Alberta Human Resources Research Council
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Curriculum Development; Curriculum Evaluation; Educational Planning; *Evaluation; Evaluation Criteria; *Evaluation Needs; *Evaluation Techniques; *Instruction; Literature Reviews; Models; Preschool Evaluation; Systems Analysis

ABSTRACT

This report brings together ideas and information about evaluation that until recently were available only to academic theorists. The four substantive chapters that follow the introduction contain (1) a general review of the "state of the art" of evaluation theory and methodology, (2) an examination of an approach to evaluation that focuses on instruction, (3) an examination of a systems approach to evaluation, and (4) a discussion of needs for the development and utilization of evaluation techniques in Alberta. (Author/JH)

ED 066788

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

D. A. MacKay

T. O. Maguire

EVALUATION OF INSTRUCTIONAL PROGRAMS

A Study Prepared

for the

Educational Planning Mission

Alberta Human Resources Research Council

EA 004 380

May, 1971

1/a

FOREWORD

This report, "Evaluation of Instructional Programs," was prepared as part of the Educational Planning Mission of the Human Resources Research Council of Alberta. Financial support was provided by the Alberta Commission on Educational Planning. The paper was initiated and completed during the tenure of Dr. Erwin Miklos as head of the Educational Planning Mission.

The authors have brought together ideas and information about evaluation which until recently were available only to academic theorists. At a time when evaluation, although much discussed, is seldom precisely defined or applied, the report performs a valuable service for teachers and administrators, who often lack both the time and the expertise to understand fully the lengthy and involved literature in the field of measurement, but who must nevertheless be knowledgeable about the role of evaluation, its strengths and weaknesses.

The opinions and views expressed in the paper are those of the authors and do not necessarily reflect the opinions and views of the Human Resources Research Council or the Commission on Educational Planning.

TABLE OF CONTENTS

		Page
FOREWORD		i
CHAPTER I	Introduction	1
CHAPTER II	The State of the Art	3
	References	17
CHAPTER III	An Example Evaluation	18
	The Description Matrix: Intents	19
	Intended Antecedents	19
	Children	19
	Teachers and Aides	21
	Facilities	21
	Intended Transactions	21
	Intended Outcomes	22
	The Description Matrix: Observations	23
	Observed Antecedents	24
	Observed Outcomes	26
	Affective Outcomes	27
	Physical Outcomes	27
	Cognitive Outcomes	28
	Processing the Descriptive Data	29
	Standards and Judgements	32
	Standards for Antecedents	33
	Children	33
	Facilities	34
	Standards for Transactions	34
	Standards for Outcomes	35
	Bases for Judgements	35
	Financial Data	35
	Judgements	37
	References	40

CHAPTER IV	A Systems Approach	40
	Introduction	40
	The Provus Approach	41
	The Four-Stage Cycle	43
	Conclusion	47
	An Example	48
	Reference	49
CHAPTER V	Projection of Needs	50
APPENDIX A	ALBERTA DEPARTMENT OF EDUCATION REQUEST FOR PROPOSALS CONCERNING A PRESCHOOL EDUCATION PILOT PROJECT IN EDMONTON AND CALGARY	52

CHAPTER I

Introduction

This final report includes the following elements:

(1) A general review of the "state of the art" of evaluation theory and methodology (Part II)

(2) An examination of an approach to evaluation which focuses upon instruction (Part III)

(3) An examination of a systems approach to evaluation (Part IV)

(4) A discussion of needs for the development and utilization of evaluation techniques in Alberta (Part V)

While this study is not a position paper as such, there are recommendations for action in Part V. The main thrust of this study is towards suggesting a discrepancy between what is being done in Alberta and what is available "out there" in the way of useful approaches to evaluating instructional programs. There are some success models available from elsewhere; the implementation of some of these in Alberta would appear to merit consideration.

It is important to point out that while the terms of reference for this study suggested that "instructional programs" were to be the object of concern, much of what we have found in our examination of the field is applicable to evaluation problems of many different kinds that are beyond the scope of a narrow interpretation of curriculum or instructional programs as such.

In general, our work has shown that there is a fairly extensive literature on this topic. Many persons in many institutions are working to develop evaluation as a special area of competence within the profession of education. It has been somewhat difficult

to encapsulate the wide-ranging endeavors in this field; but it now appears that the field is settling down and that consensus is being achieved as to what the priorities ought to be in educational evaluation and how best it can be carried out. In many ways, evaluation as a special activity of a fairly sophisticated type is, of itself, an innovation in educational organizations. There are deficiencies to be accounted for and, most difficult of all, walls of resistance to be broken down. The current use of slogans such as "accountability" and the piece-meal adoption of some of the tools of evaluation and planning (e.g. cost analysis) without a broad conceptual framework may have retarded progress to a degree. However, a fairly objective, albeit somewhat cursory, review of the possibilities may be of service to those who are engaged in educational planning in Alberta.

CHAPTER II

The State of the Art

The variety and complexity of problems that have faced mankind since the onset of the atomic age need no documentation here. It is sufficient to note that the extinction of the species is now a distinct possibility. Despite the gloomy predictions of many, societies continue to struggle for their survival. As each society faces its problems, it seems to develop characteristic modes of attack. Often these modes take the form of investing the responsibility for solution with a particular social institution. In postwar English speaking societies, the social institution called upon to solve many of the problems, has been Education.

Stated in the extreme, the panacea of the 50's and 60's has been that education in some form is capable of solving any problem. This faith, as is often the case, has been carried the furthest in the United States. For example, the scientific "crisis" created by Sputnik was treated with massive aid to education in an effort to upgrade the scientific qualifications of American youth. Or, more recently, the problem of racial inequalities was treated with a variety of compensatory education schemes.

In Canada, the situation is similar though less pervasive. As the Quebec "problem" was brought to the consciousness of English society, more French language education was provided, student exchanges were conducted, and bilingual schools implemented. Similarily,

unemployment and unequal economic opportunities across Canada stimulated large federal assistance to vocational education which resulted in the construction of dozens of vocational high schools. Other examples exist.

It is not the purpose of the writers to debate the appropriateness of the demands placed on education. The demands have been numerous, urgent, difficult, and often conflicting. In many cases the responses to these demands have required the investment of enormous amounts of national resources. So large have been the costs that other important goals have been neglected. It is not surprising then that society has begun to ask for an accounting from the educational establisiment. As a result, many kinds of questions are being asked. Are the provided solutions valid? Have they been worth the effort and cost? Are the suggested ways of dealing with a problem better than other ways? What are the effects of the solutions on the students, parents and teachers? These are only some of the questions that exemplify the public's desire for evaluation of the efforts of the education institution.

One of the responses by the educational establishment has been a loose knit set of models, recipes and practices which are grouped into a technology known as curriculum evaluation. Curriculum evaluation broadly defined refers to the determination of the merit of an instructional program. There are many ways to structure the field. One could classify the methodologies by their disciplinary biases, such as psychometric, economic, psychological, sociological, etc. One could look at whether the evaluation is focused on process

or product. One could even try to identify separate schools of evaluation thought. Each classification has some use in relating the various models, recipes and practices, and in the review which follows the various dimensions will be used. However, for those who are unfamiliar with the field, it is probably least confusing to look at its development in a fairly chronological fashion. Therefore in an effort to provide a flavor of the diversity of tactics and procedures that have been employed within the total field of curriculum evaluation, a review of the emergence of curriculum evaluation as a technology in the United States will be undertaken. Following the review, a more detailed description of two models will show how evaluation can be applied in specific circumstances.

It seems presumptuous to try to pin down the exact source of current evaluation thought, but most educators would have to agree that the work of Tyler, especially examplified by the activities of the evaluators in the eight year study (Smith and Tyler, 1942) was an important early milestone in the development of a technology of curriculum in terms of how well observed behaviours matched the stated objectives.

Tyler's model has several merits. It provides valid, reliable and objective data for an evaluation. It makes differentiated evaluation a possibility by allowing the evaluator to indicate which objectives were achieved and which were not. In addition, the behavioral statement of objectives is likely to make both curriculum development and teaching become more systematic.

On the other hand, strict application of the Tyler model had some difficulties. The statement of objectives in behavioral terms is a long and often tedious procedure. In addition, it is too easy to avoid questions about the worth of the objectives themselves, particularly if the evaluation is carried out after the objectives are set. A further criticism of the model lies in its restriction to those objectives specified in advance. No systematic search is conducted for other resulting behaviors. Finally, the often legitimate question of comparison of one method or curriculum with another is explicitly avoided.

In some respects, the work of Tyler seems like a reaction to the tradition of educational research. From the time of E. L. Thorndike, one of the principal tools of researchers has been the comparative experiment. In its most basic form the comparative experiment involves a comparison between two "equal" groups, one of which received a treatment. As a research tool in education, the comparative experiment predates Tyler's work by several decades. However, its use as a valid evaluation model became most prominent in the mid fifties. Prior to this time, attempts were made to compare curricula, but the validity of the experiments is so suspect that in general the results cannot be treated seriously. In the field of experimental design, the systematic development and widespread dissemination of valid procedures for comparative studies is a relatively recent thing.

In the mid fifties the rush into curriculum development prompted by the competition between the U.S.A. and Russia for space supremacy

produced a strong thrust for valid comparative evaluations. It was assumed that old methods of learning and instruction were no longer adequate and must be improved or new ones would have to be found. This produced a natural question of comparison between new and old curricula. The work by Fisher (1945), Lindquist (1953) and much later by Campbell and Stanley (1963) provided some of the background expertise to curriculum evaluators so that valid comparisons could be drawn.

While the comparative evaluation provides the direct answer to a simple evaluation question (which is better?), the requirements for validity are stringent. All too often comparisons were made between programs serving different populations. The definitive answers that were promised could seldom be produced. The conclusions of many evaluations were very often that students in the traditional curriculum performed better on tests measuring traditional goals than students in the new curriculum, and students in the new curriculum performed better on tests measuring the new goals.

For ten years or so comparative evaluation studies held sway. Then in 1963 Cronbach raised a number of questions concerning the usefulness of their role in course improvement. Cronbach noted that evaluation was used in the service of course improvement for deciding what instructional materials and methods are satisfactory and for deciding where change is needed. He pointed out that evaluation should not only show what the effects of a curriculum are, but also it should show how the effects are achieved. He indicated that

global comparative studies were rarely definitive enough to justify the expense involved and advocated the use of several sources of evidence such as process studies, proficiency measures, attitude measures and follow-up studies. By using a variety of instruments the unanticipated outcomes of a curriculum could be detected.

At about the same time and in the same vein as Cronbach's suggestions, Taba and Sawin (1962) proposed a model of evaluation which focused on the collection of information which would determine why some students failed to achieve stated objectives. Some of the evidences to be collected included observations on teaching method, patterns of classroom interaction, physical facilities, and student abilities and motivations. Together with Cronbach's work, Taba and Sawin's ideas were a major shift in the focus of evaluation from the outcomes of learning to the process of learning. Their emphasis was on evaluation in the service of curricular improvement.

Continuing in the same direction as Cronbach, Walbesser (AAAS Commission on Science Education, 1965), retained some of the most useful elements of the Tylerian model and bent them to the evaluation for course improvement. Walbesser referred to principles of Gagné's hierarchies of objectives when he suggested that course objectives be broken down into prerequisite objectives. He noted that if objectives are organized into hierarchies, then the achievement of an objective at one level is dependent upon the achievement of constituent objectives at a lower level. Consequently when learning difficulties occur in a curriculum, the deficient portion of the

curriculum can be pinpointed rather exactly. In a way this procedure amounts to the use of the Tyler model at a micro level.

Although "Neo-Tylerian" models such as Taba and Sawin's and Walbesser's have useful qualities for the curriculum developer, they are found wanting from the consumer's view. Their major shortcoming was that they neglected the whole dimension of value. Objectives and content are not the only characteristics that are pertinent to curricular decisions; costs, effects on teacher workloads, ease of implementation, social importance and appropriateness of teaching method are only a few of the value laden variables that are important for teachers, school boards and the public to know about.

In an effort to cover some of the inadequacies of the existing models, Taylor and Maguire (1966) proposed a framework for evaluation which was based on a four stage conception of curriculum development. They suggested that the needs of society are interpreted by various social agents into broad educational goals. Curriculum developers translate the broad goals into more specific behavioral goals and then develop classroom strategies to attain them. The students interact with the strategies to produce observable behaviors. Evaluation in general consisted of two kinds of activities: measuring and assessing value. The measurement component was seen to consist of the description of goals, environment, personnel, methods and outcomes as well as the determination of the relationships among them. The value component included the collection of judgements of quality and appropriateness of the goals, strategies and outcomes. At

each stage, and between stages the descriptions and judgements were compared and combined to produce a profile of strengths and weaknesses.

The framework developed by Taylor and Maguire differed from earlier work by incorporating the assessment of value at various points in the evaluation process. However it did not go as far as presenting a broad conceptualization of the methodology of evaluation. This was provided by Scriven.

Although the paper entitled "The Methodology of Evaluation" by Scriven was not published until 1967, it was circulated in mimeograph version two or three years earlier. It has probably been the greatest single influence on the field of curriculum evaluation. Prior to its circulation, writers in the field were bogged down in a mire of semantic confusion. Evaluation as a term meant something different to every writer. Scriven's contribution was to set the evaluation house in order.

Scriven noted that the distinction between the roles of evaluation and the goals of evaluation is blurred, very often intentionally. Evaluation plays a role in curriculum development, in decision making, in course improvement and elsewhere, but whatever its role, the goals are always the same - to estimate the merit, worth, or value of the thing being evaluated. Scriven pointed out that the subversion of goals to roles was very often a misguided attempt to overcome the anxiety in those educators whose products and activities are being evaluated. The consequence of this kind of mutiliated evaluation could be much more undesirable than the anxieties evoked.

A second clarifying distinction made by Scriven was the distinction between formative and summative evaluation. These labels refer to two broad roles of evaluation. The term formative evaluation refers to the evaluation of courses when they are in a state of development. Summative evaluation refers to the assessment of curricula that are ready for the market. This distinction has implications for the personnel involved in the evaluation. The formative evaluator must work in close cooperation with the curriculum director. For the summative evaluator quite the opposite is true. He must be free of any potential stigma of conflict of interest so that his evaluation has an itegrity of design and conclusion.

Scriven also took issue with Cronbach on the role of comparative studies. While agreeing that comparative studies are very often equivocal or else do not give any understanding of why observed differences exist, Scriven suggested that comparative evaluations are often easier than absolute evaluation, and that the results of a comparative study are useful at various times in the development of a curriculum to provide the global information needed to decide whether to continue with development or scrap the program.

Scriven's contribution to the theory of evaluation was monumental, but it did not provide many of the answers to the "nuts-and-bolts" kinds of questions that the practicing evaluator was forced to deal with. For workers in the field, there was a need to spell out rather explicitly the procedures and instruments necessary to carry out a valid evaluation. This need was met by a number of writers, among whom Stake, Stufflebeam, Alkin and Provus are important examples.

Stake focused on the data of evaluation, noting that in general it could be divided along two dimensions. One dimension separates the data into descriptions and judgements. On the other dimension, data are classed as antecedent, transaction or outcome. Antecedent data are descriptions and judgements collected on conditions prior to the program. Transactions are descriptions and judgements of activities that occur as the program is carried out, and outcome data refer to the results of the program. Having classified the data, Stake showed that for him, evaluation consisted of determining the degree of relationship and agreement among the various classes of the data.

Stake's model is perhaps the prime example of distinct school of evaluation thought. Primarily psychological and psychometric in training its adherents have stressed the need for understanding what Hastings (1966) has called the "Whys" of educational outcomes. The data collection net is cast widely so as not to miss any possible variables which might be relevant to the relationships among antecedents, transactions and outcomes.

Both the strengths and the weaknesses of this model lie in its lack of disciplinary blinders. On the one hand because of the broad base laid for data collection, possible relationships stand less chance of being missed than they do in models which use a theoretical framework for determining which data to collect (for example Walbesser's model). On the other hand because of its scope and the finite resources of evaluations, important relationships may not be investigated as thoroughly. Proponents of Stake's model would

describe it as an all inclusive model. Critics might label it blindly empirical.

In contrast to the intentional vagueness of Stake's model are a number of models that have developed from the role that evaluation can play in educational administration. The focus of data collection for these models has been on variables that are necessary for arriving at specific curricular decisions. The principal example of the development of an evaluation model based on a decision making rationale is the work of Stufflebeam (1967).

In Stufflebeam's model, evaluation is defined as the process of acquiring and using information for making decisions associated with planning, programming implementing and recycling program activities. His model has been called the CIPP model after the four stages of evaluation that he describes. In the first stage, Context Evaluation, the goal is to identify and assess needs and to identify probelms underlying the needs. The second stage is Input Evaluation in which the evaluator assesses system capabilities, available input strategies, and designs for implementing the strategies. In Process Evaluation, the goal is to identify and predict in process, the defects in the design or its implementation. The final stage of evaluation is Product Evaluation in which the goal is to relate outcomes to objectives and to context, input and process information.

Each of the four stages of evaluation is related to a decision making process. Context evaluation is useful for deciding upon the

setting to be served and the goals to be sought. Input evaluation is used for selecting sources of support, kinds of strategies to be used for problem solution, and procedural design. Process evaluation is useful for implementing and refining the program, and of course outcome evaluation is necessary to decide whether to continue, modify or scrub the program.

Alkin (1970) and the Staff of the Center for the Study of Evaluation (CSE) at UCLA have followed in the paths blazed by Stufflebeam and his associates at Ohio State in attempts to associate the tasks of evaluation with the responsibilities that decision makers have relative to educational programs. The CSE definition of evaluation is "the process of ascertaining the decision areas on concern, selecting appropriate information and collecting and analysing information in order to report summary data useful to decision makers in selecting among alternatives". In short, evaluation plays a role primarily as an adjunct to decision making.

Five decision areas have been listed as important to the improvement of instruction. They are: problem selection, program selection, program operationalization, program improvement and program certification. Corresponding to the decision areas are five evaluation requirements: Needs Assessments, Program Planning, Implementation Evaluation, Progress Evaluation, and Outcome Evaluation. Needs assessment attempts to examine the gap between specific goals and existing situations. In program planning evaluation the evaluator trys to assess a program's potential for success. The task of implementation evaluation is to collect information on how well the program is being

implemented. Progress evaluation is a functive evaluation for program modification. Outcome evaluation is similar to summative evaluation in as much as it is related to program certification.

Following their conceptions of evaluation CSE has produced an Elementary School Evaluation Kit which can be used to help administrators evaluate their elementary schools. At the present time, the kit is focused in needs evaluation. A proposal to study the application of this material to the Western Canada situation is under consideration at the present time in the Educational Studies Areas of HRRC.

A more definitive application of systems methodology was provided by Provus (1969). He presented a model which arose from attempts to combine evaluation technology with management theory for the evaluation of curriculum innovations within a large school system. Provus noted that evaluation essentially consists of (a) agreeing upon program standards, (b) determining whether a discrepency exists between the standards and the program, and (c) using the discrepency information to correct weaknesses in the program. Four stages of evaluation corresponding to four stages of program development were defined: Definition, Installation, Process and Product. The process of evaluation consists of moving through the four stages and through three major content categories: Inputs, Processes, and Outcomes.

In many respects the model is very much like Stake's model cast along a curriculum development continuum. The content categories and their subdivisions correspond rather closely to Stake's Antecedents, Transactions and Outcomes. The first stage,

definition, is very similar to what Stake has called intents. The work of the evaluator in the installation stage of the Provus model is similar to the work of the evaluator as he observes the transactions. At the process stage the evaluator responds to the observed outcomes and tries to relate them to the transactions. At the product stage, the evaluator looks for congruencies between intents and outcomes.

This capability of being able to map (at least superficially) the Provus categories onto the Stake matrices should not be taken as casting doubt on the value of Provus' model. To some extent the same interrelationships exist among all models of evaluation. Provus' model has many merits, not the least of which is the specificity with which the procedures are spelled out. Such explicitness makes a valid model very useful to the naive evaluator.

In summary, it is useful to highlight three kinds of models of evaluation. Each reflects the background and concerns of its authors. Neo-Tylerian models such as Walbesser's focus on the learning process and the sequences of objectives necessary for achievement. The role of evaluation is primarily fomative. The eclectic models, like Stake's, focus on the collection of data both to answer and to raise questions. Administrative models, like Stufflebeam's, are closely tied to the collection of information for particular decisions.

The range and intensity of activities suggested by the models may vary greatly but the commonalities are compelling. In all cases the role of objectives is prominent. In all cases it is recognized that poor results are often due to the slip between the cup of intention and the lip of practice. And, in all cases, the ultimate determination of worth lies in human judgement.

References

AAAS Commission on Science Education An Evaluation Model and its Application. Washington, D. C.: American Association for the Advancement of Science. Publication 65/9, 1965

Aikin, M. C. Evaluation Theory Development. Evaluation Comment 2:1, 2-7 1969. Los Angeles: Center for the Study of Evaluation, University of California.

Campbell & Stanley Experimental and Quasi Designs for Research on Teaching in N.L. Gage (ed) Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

Cronbach, L. J. Course improvement through evaluation. Teachers College Record 1963, 64, 672-683

Fisher, R. A. The Design of Experiments London: Oliver and Boyd, 1945

Hastings, J. T. Curriculum evaluation: the whys of the outcomes. Journal of Educational Measurement 1966, 3, 27-32.

Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. Houghton-Mifflin, 1953.

Provus, M. Evaluation of Ongoing Programs in the Public School System. Educational Evaluation: New Roles,New Means, in Tyler, R. W. (ed) Sixty-eighth Yearbook of the National Society for the Study of Education Part II, University of Chicago, Press 1969.

Scriven, M. The methodology of evaluation in Stake, R. E. (ed) AERA Monograph Series in Education, No. 1 Chicago Rand McNally, 1967, 39-89.

Smith, E. R. & Tyler, R. W. Appraising and Recording Student Progress. New York: Harper and Row. 1942.

Stake, R. E. The countenance of educational evaluation. Teachers College Record 1967, 68, 523-540.

Stufflebeam, D. L. The Use and Abuse of Evaluation in Title (T) Theory Into Practice 1967, 6(3), 126-33

Taba, H. and Sawin E. A proposed Model in Evaluation. Educational Leadership 1962, 20, 57-71

Taylor, P. A. and Maquire, T. O. A theoretical evaluation model. The Manitoba Journal of Education Research, 1966, 1, 12-17

CHAPTER III

An Example Evaluation

In the first section of the present report, a review of evaluation methodology was presented. In the present section an attempt will be made to show how the methods described previously might be applied to a program in the province of Alberta. For this purpose, it was decided to select a curriculum area whose evaluation is of some relevence to the contemporary scene. Such an area is preschool education.

In 1970, the Department of Education of Alberta requested proposals from interested parties concerning the establishment of pilot preschool education projects in Edmonton and Calgary. The objectives of the program were listed in a memo issued by the Minister of Education. After adjudication, two proposals for pilot projects were funded by the Department of Education, one in Edmonton and one in Calgary. In the present chapter an example evaluation outline will be proposed that could be adjusted for use in the Edmonton project. The outline will be based on two documents.

1. The Request for Proposal Issued by R. C. Clark, Minister of Education (RFP).
2. Edmonton Preschool Education Pilot Project - Detailed Submission from Edmonton Public School Board (Proposal).

Because the actual activities subsequently undertaken by the recipients of the grants may have been revised from those described in the above documents, the example evaluation outline should not be taken as valid for the project as it exists. Rather it should be taken

more as a hypothetical evaluation from which appropriate elements could be selected for use in the project as it was actually carried out.

For the purpose of developing an evaluation outline, Stake's (1967) model will be used as a basis, but the work of other methodologists will be incorporated as needed.

A representation of the data to be collected is shown in Figure 1. Each of the cells will be considered in turn and the source and kinds of data to be collected will be specified.

The Description Matrix: Intents

Intended Antecedents

In this example, the intended antecedents refer to the kinds of children that the program is intended to serve as well as the facilities and teachers to be used. These factors are spelled out in the RFP and in the Proposal and can be divided into three sections. Some examples of intended antecedents are given below.

Children

1. Children come from inner city core.
2. Children must be eligible to enter grade one in the following year.
3. Children must be culturally handicapped. Culturally handicapped includes:
   - children from low income homes
   - children from broken homes
   - children who receive little love or attention
   - children whose parents speak a minority language only
   - children who lack experience working and playing with others

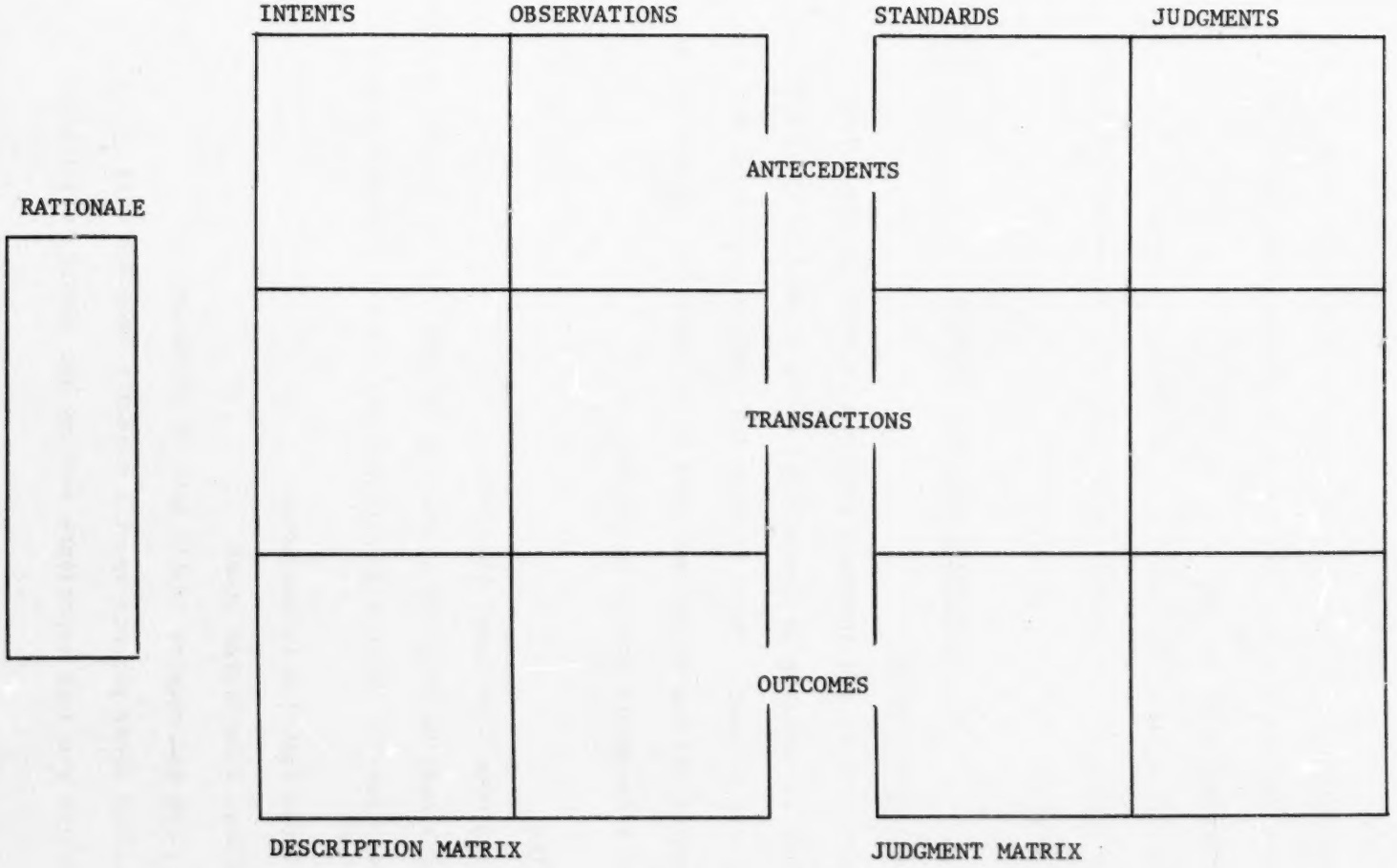

Figure 1. A layout of statements and data to be collected by the evaluator of an educational program.

4. Children should form a heterogeneous group with respect to race, religion and background of parents.

Teachers and Aides

1. There will be four teachers with undergraduate training in preschool education. The teachers will be capable of supplying a warm responsive climate.
2. There will be four teaching aides with the following characteristics;
   - knowledge and appreciation of the learning and teaching process in the school.
   - ability to work under teacher supervision and to relate to students.
   - able to play the piano and sing.

Facilities

1. Furniture: chalkboard, sand table, study chairs, trapezoidal tables, etc.
2. Music: 5 pr rhythm sticks, 4 jungle clogs, 1 piano, etc.
3. Records: Modern Mother Goose, Adventures in Music I, etc.
4. Science: wires, batteries, magnifiers, etc.
5. Physical Education: wagon, hoops, 6 bean bags, 2 balance boards, etc.
6. Manipulative toys

Intended Transactions

The intended transactions are the procedures that are to be used to produce the outcomes of the program. Again, these intentions can be collected at a general level from the Proposal. Of course, the Proposal does not give a detailed specification of the curriculum since part of the project involves curriculum innovation, but the general guidelines

are laid out, and several examples of intended transactions are listed below.

1. Assess the social, physical and emotional needs of children when they first enter the school.
2. Provide an environment and experiences appropriate to these needs.
3. Provide a climate of trust warmth and security.
4. Involve the child's parents in the assessment of needs and in the determination of appropriate learning experiences.
5. Use a program which incorporates:
   - physical activity
   - motor perceptual activity
   - discussion with adults
   - experience with books, and suitable math, science and language materials.
   - creative expression through art, music and rhythms
   - appropriate routines and regulations

Specific transactions are listed for psychomotor development, motivational development, attitude development, and cognitive development. For psychomotor development, the following specific activities are listed:

- outdoor exercises such as running, jumping, climbing and digging
- indoor play with blocks, and toys
- art experiences such as modeling, drawing, and painting
- work with blocks, puzzles, balls, hoops and bean bags.

Intended Outcomes

In the preschool project the intended outcomes are closely tied to the intended transactions. Although they are not specified in behavioral

terms, the level of generality of the outcomes that are listed is sufficient for use in the present example. There appear to be two levels of outcomes. The most general outcome is stated in the RFP as:

"The aim of this program will be to enable each child to adapt successfully to the demands and opportunities of elementary school life."

The more specific outcomes are listed in the Proposal. Some examples are shown below.

1. To develop in each child an attitude towards himself and others that is conducive to a positive self concept.
2. To help the parents widen and enrich their knowledge and understanding of their children.
3. To develop large muscle groups.
4. To improve hand-eye and fine muscle coordination.
5. To develop patterns of satisfactory group living.
6. To develop a spirit of exploration, experimentation and creation.
7. To develop the ability to describe, explain and inquire effectively.
8. To develop a base for learning by developing through physical, active, sensory, concrete and manipulative stages, towards verbal, symbolic and abstract stages.
9. To develop an ability to ask questions, classify information, draw conclusions, and make inferences.

The Description Matrix: Observations

In the first three cells of Stake's model, the intentions of the program are spelled out. In the second group of cells, we turn to the

project as it actually goes on. Again, we can focus on antecedents, transactions and outcomes, and try to devise ways of describing what occurs. This activity is guided by the elements that have been listed in the Intentions column, but an attempt is made to go beyond the outcomes specified in order to pick up any unanticipated effects that may result. In the following section, some procedures and instruments will be listed that would be useful for describing the program. No attempt will be made to provide a complete list, but extensive examples will be given to provide a flavor of what the evaluation might look like.

Observed Antecedents

A description of the children's home environment could be undertaken using Mosychuk's (1969) DEPVAR Scale. This scale which uses an interview format provides scores on ten Environmental Process Variables. The variables that are measured are:

1. Academic and Vocational Aspirations and Expectations of Parents
2. Knowledge of, and Interest in Child's Academic and Intellectual Development
3. Material and Organizational Opportunities for the Use and Development of Language.
4. Quality of Language in the Home
5. Female Dominance in Child Rearing
6. Planfullness, Purposefullness and Harmony in the Home
7. Dependency Fostering -- Overprotection
8. Authoritarian Home
9. Interaction with Physical Environment (Visual and Kineasthetic Experiences)

10. Opportunity for, and Emphasis on, Initiating and Carrying Through Tasks

These scales were found to be related to intellectual development of various kinds. They appear to measure a degree of cultural deprivation. The measurement would be made at the beginning of the school year, by interviewing mothers of children enrolled in the preschool classes.

Administration of the Weschler Intelligence Scale for children would provide an assessment of several kinds of mental abilities.

The teacher and teacher aides would also be interviewed in order to assess their qualifications in relation to the intended qualifications. In addition, the Carkhuff Scale would be administered to measure level of communication.

In order to determine the facilities actually present for use in the program, a mid year inventory would be taken of all equipment and materials.

Observed Transactions

Two kinds of transactions were suggested in the intents column. The first, more general set of transactions referred to climate or environment. The second type referred to more specific kinds of classroom procedures. In an effort to describe what occurs, four kinds of data can be collected.

1. Observation schedules. Periodic observation of the classroom situation can be undertaken to describe the amount and variation of time spend on various kinds of physical activity, or various kinds of intellectual activity. The amount of interaction with adults can be determined as can be the amount of time that each child spends on individual or group activity.

2. Teacher interview. One of the transactions that is intended is the establishment of a warm climate. Statements can be taken from teachers indicating what steps they are taking to provide such a climate.

3. Student Interviews. The students can provide valuable insights into the climate of the classroom. One method that is useful is to use the My Class Instrument development by Anderson (1971) as a focus for the interview. Another although somewhat more difficult procedure is to have individual children draw a picture of the class and then ask them about the picture.

4. Parent Interviews. One of the intended transactions is to involve the parents in the program. A sample of parents can be interviewed to determine their impressions of the program, and to see if the childrens' home behavior reflects a climate of trust warmth and security at school.

Observed Outcomes

Most of the outcomes listed in the intents column were expressed in terms of "to develop", "to change", etc. As a consequence it would be necessary to measure the outcomes on a pre-test post-test basis. This poses some problems with appropriateness of the measures, for children who could be as young as four and one half at the time of initial testing. In selecting instruments, that consideration must be kept in mind. A second consideration is that there are two kinds of outcomes, the immediate (end of year) outcomes and the long range outcomes (after experience in the public schools). It might be useful to separate these and discuss the immediate outcomes first.

The immediate outcomes are classifiable into Affective, Physical, and Cognitive. In the following paragraphs, several instruments will be suggested. They are meant to be guides, rather than prescriptions. The instruments will be appropriate to a greater or lesser extent depending on how the program is actually implemented.

Affective Outcomes. These outcomes can best be measured using observation schedules. The Pupil Behavior Inventory of Winter, Sarri, Vornwaller, and Schafer (1966) consists of the following eight scales that seem appropriate to the objectives that were listed in the intended outcomes.

1. Dependence
2. Inner controls
3. Interaction with other children
4. Ability to get along with other children
5. Comfort in school
6. Achievement Motivation and pride of Mastery
7. Curiosity
8. Creativity

The inventory has generally been used by having teachers rate the children on the items that compose the scales. In this case it would be better to develop the Inventory into an observation schedule. This would overcome problems in reliability.

Physical Outcomes. The Purdue Perceptual Motor Survey (Roach and Kephert, 1966) contains items that are useful measures of gross motor coordination. The Frostig Developmental test of Visual Perception

(Frostig, 1963) would be useful to pattern measures of fine coordination after.

Cognitive Outcomes. Few standardized tests exist that would adequately measure the outcomes specified in the proposal. Nevertheless, a number of instruments have been developed that measure some of the cognitive objectives. In addition, these instruments provide good examples of items that can be used for children of the preschool age level and tests could be tailored for the specifics of the project using the same kinds of items. In the following list, some of the available preschool tests are presented. In some cases subtest scores are possible, and these are shown as well.

1. Moss Test of Basic Information - Moss, (1967)
2. Preschool Inventory - Caldwell (1967)
   - Personal-Social Responsiveness
   - Associative Vocabulary
   - Concept Activation-Numerical
   - Concept Activation - Sensory
3. Basic Concept Inventory - Englemann (1967)
   - Basic Concepts
   - Statement Repetition and Comprehension
   - Pattern Awareness
4. Preschool Academic Skills Test - Provus, Kresh and Green (1968)
   - Verbal Labeling
   - Color Labeling
   - Classification

Functional Relationships

Visual Matching

Auditory Matching

Picture Arrangement

Symbol Series

Counting

Verbal Concepts

5. Piaget Procedures of Summative Evaluation - Kamii (1971)

The long range goal, that is, adaptability to school might be looked at in a number of ways. Judges who were unfamiliar with the children could be asked to observe the children after they had been in the first grade for six months and try to sort all of the children in the grade one classes into two piles, those with preschool experience and those without. Within any class, there will likely be enough children without the experience to provide the basis for a valid test.

A second, and more common procedure would be to administer some standardized beginners' tests to the preschool children and compare the results with results from the area in other years.

Processing the Descriptive Data

Prior to considering the judgement matrix, it is useful to consider the steps that would be taken in processing the descriptive data. As Stake notes, there are two principal ways of processing descriptive data: finding the contingencies among antecedents, transactions and outcomes, and finding the congruences between intents and observations. The format for processing the data is shown in Figure 2.

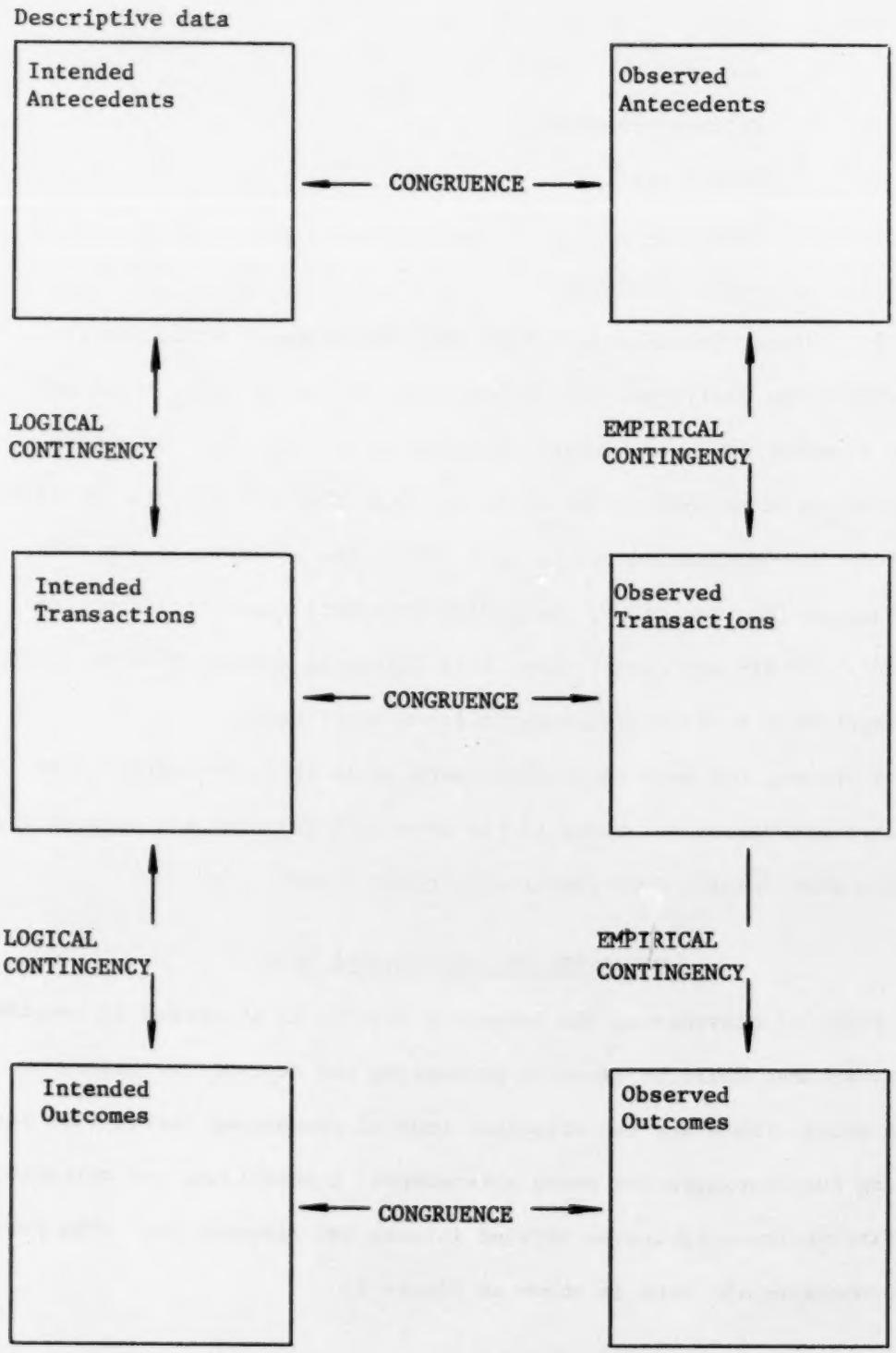

Figure 2. A representation of the processing of descriptive data.

Stake points out that to be fully congruent, the intended antecedents transactions and outcomes would have to come to pass. Two points should be made. Firstly, it may be that incongruence is desirable in the long run, especially, if the intents can be shown to be invalid for some reason. Secondly, congruence does not assure validity, only fidelity.

The degree to which the observations match the intents becomes a question for the standards and judgement procedures to be discussed later. At this point we are looking for a qualitative match.

Contingency establishment may be of two types, logical and empirical. Logical contingencies refer to the relationships that should exist between intended antecedents, transactions and outcomes. We ask the questions of the following sort, "If we use teachers with certain qualifications, and have students with certain backgrounds, and then we apply certain methods is it reasonable to expect the outcomes listed in the intended outcomes cell?" Questions of this type could be answered by experts in early childhood education, developmental psychology and learning psychology. In short, the establishment of logical contingencies is a question of expert judgement. In the present example, we could ask our assembled group of experts to consider the three intent cells and judge the extent of logical contingency among them.

Empirical contingencies provide data for the following sorts of statements. When teachers provide independent children with the opportunity to explore, with blocks, the children tend to learn to conserve, When teachers provide dependent children with a structured matching task they tend to learn to conserve. These kinds of statements are based on the observations made. In many cases they are statistical problems of estimating the relationships among variables.

The establishment of congruences and contingencies is important for course revision. Broken congruences, or poor contingencies often point to flaws in the curriculum. Careful analysis makes it possible to establish the point of breakdown in the curriculum. For example, if it were discovered that the children were unable to attach labels to certain objects, it might be possible to trace the problem back to a lack of congruence between an intended antecedent that the children would have had a certain kind of home experience, and an observed antecedent that they did not have the antecedent experience, and that without the experience, the subsequent teaching strategy could not possibly succeed in inducing the desired outcome.

Standards and Judgements

Stake's model incorporates as part of the basic data a Judgement Matrix that is composed of two parts, standards and judgements. For most situations there will be no ready made sets of standards to apply to the discriptions from the Description Matrix. More often than not, standards must be created by the evaluator. Obviously there can be as many sets of standards as there are interested parties. The task of the evaluator is to collect the appropriate sets.

The purpose of judgement is to weigh the importance of various standards, to measure the intents and observations against the significant standards, and to combine the measures into a useful evaluation of the merit of the program.

Without having a detailed knowledge of the preschool program both as it was intended and as it occurred, it is difficult to suggest useful standards that could be applied. Nevertheless, an attempt will be made in

the following discussion to provide some examples of standards that might provide useful examples for stimulating more appropriate standards.

Standards for Antecedents

There are three components in the antecedent cells and we can consider them separately.

Children. On the surface, it is difficult to conceive of standards for children, but if we consider the specifications for the intended student clientele, we note that the program was set up for use with culturally deprived children. It will be useful to consider some standards for cultural deprivation. Three sources come to mind. The Mosychuk (1969) study provides data indicating expected scores on his scales for a working class neighbourhood in the city of Edmonton. Average income level could be obtained from census data for various tracts in the city, and sociologists could be called upon to rate the degree of cultural deprivation in the children who attend the preschool classes. These three sources would provide some standards against which the sample could be measured, to determine whether or not the students who are in the program match the expectations laid down in the RFP.

Teachers. Two kinds of standards for teachers seem useful. The first relates to their academic qualifications. A statement of the qualifications of teachers employed in other preschool situations in Alberta would provide one set of standards for academic qualifications, another set could likely be obtained by reviewing the preschool education literature.

A second kind of standard refers to what might be called the "human qualifications" of the teachers. The Carkhuff (1969) scale provides a

measure of how well people are able to communicate with other people. The scale has built into it an implicit set of standards along which communication skill can be assessed.

Facilities. Two sources of standards for facilities seem useful. A group of preschool education experts could be assembled to draw up a list of necessary and desireable items for use in preschool program, and the existing list could be compared with the ideal. A second source could be compiled by searching the literature for lists of equipment used in other programs.

Standards for Transactions

One of the most contentious issues in education is the determination of standards for instruction. In an effort to set standards for the present situation, experts representing various pedagogical points of view (Piagetian, Montessori, behavior modification, Dewey, experiential, etc.) could be brought together and allowed to view videotapes of classroom transactions. The experts would then be asked to write a critical analysis of what they saw, relating their analysis to the dictates of their pedagogical philosophies.

In other curriculum projects, the standards for transactions might be less difficult to come by. The Flanders Interaction analysis has become so widely used as one kind of transaction measure that it is now possible to get some normative information for a variety of situations. For example, the research indicates that a certain ratio of direct to indirect teaching is commonly found in certain kinds of classrooms. This kind of standard would be useful in situations where indirect teaching was one of the intended transactions.

Standards for Outcomes

Several of the measures listed on the Observed Outcomes section provide percentile norms for interpreting the scores that are observed. Of course, caution must be observed in interpretation as most of the norms are not Canadian, nor will they be based on extensive samples. Nevertheless, some indication of standards is possible.

Bases for Judgements

Stake notes that there are two bases for judging a program; judging with respect to absolute standards and judging with respect to relative standards as characterized by alternative programs. He symbolizes this process in Figure 3.

For the present project, several standards have been suggested against which the descriptive data can be compared. The task of judgement is to decide what levels are to be considered sufficient. More attention will be paid to this problem at the conclusion of this chapter.

The second basis for judgement can be satisfied by making comparisons with other programs. The entire description matrix could be collected from the Calgary project, as well as from preschools run by Edmonton Kindergartens Ltd. It is important that the entire matrix be collected because, the various preschools will have differing emphases and differing clientele. Further comparisons on outcome variables could be made with children from similar backgrounds who receive no preschool education.

Financial Data

Prior to making the final judgements about the program, it is necessary to attend to the financial factors. Stake makes no explicit reference to finance evaluation, although it is implicit in all stages, (for example,

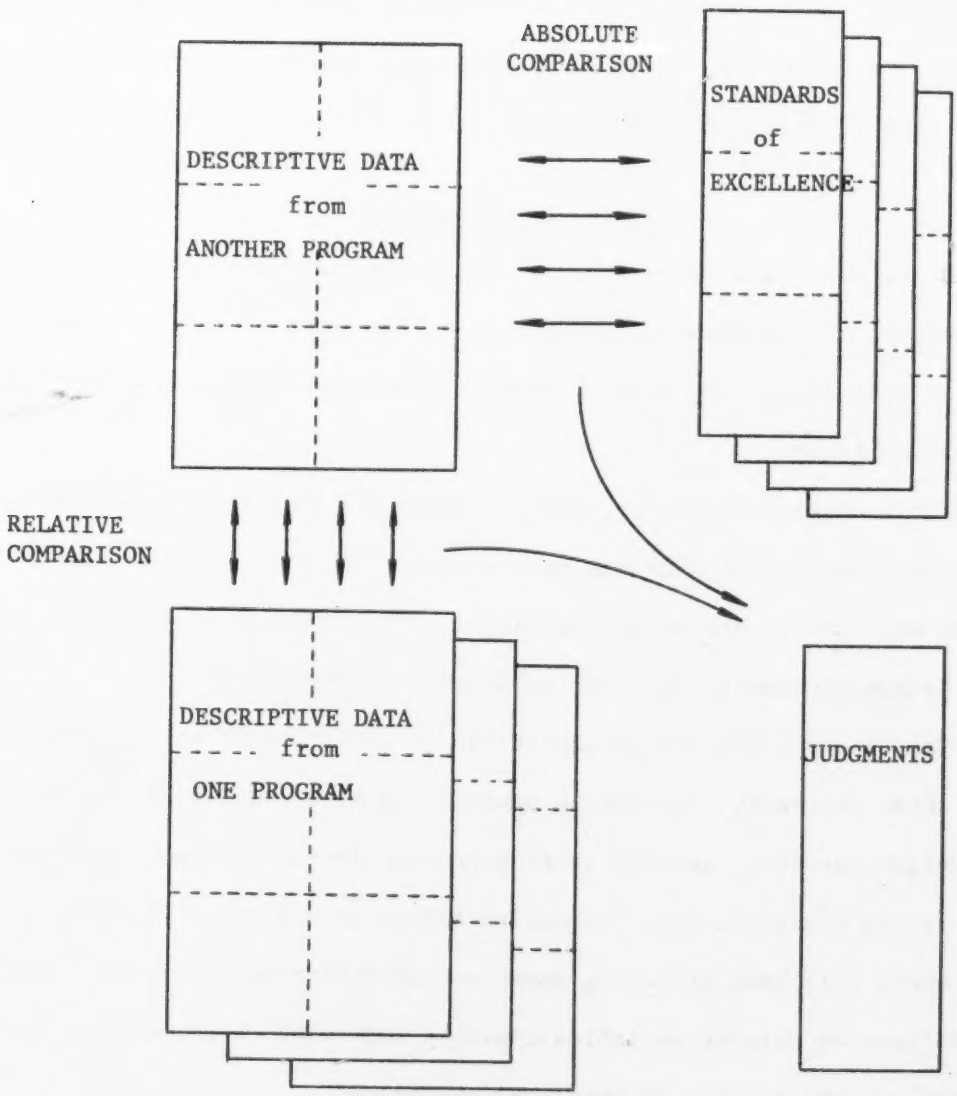

Figure 3. A representation of the process of judging the result of an educational program.

an intended outcome might be to have the most economical program possible). In the evaluation of the preschool education pilot project, the financial factors must be developed explicitly. For this purpose, accountants could be hired to set out a classification of costs in the program. This activity could be carried out in the alternative preschools as well.

In order to establish some absolute standards concerning the costs, a sample of civic taxpayers could be taken in which respondents would be asked to indicate whether they would be in favor of raising their taxes to cover the costs of the preschool education program. The approximate dollar increase in taxes that would be favored by various percentages of the population would provide the standard for judgement. For example, it might be that 90% would accept a $10.00 increase per year, 50% would accept a $20.00 increase, and 10% would accept a $50.00 increase. Such a scale would be useful for the deliberations described in the next section.

Judgements

Stake suggests that one of the tasks of the evaluator may be to judge the program. In the present case, since the decision lies with the various officials of the education establishment, it would be wasting effort for the evaluator to do the judging. One useful possibility would be to select a blue ribbon committee to sit as a board of judgement. Members of the committee might include ranking members of the Department of Education, school board representatives, and taxpayers (including parents of children in the program). A report of the evaluation would be distributed to the committee so that the members would be familiar with its contents prior to meeting. At the meeting, evaluators would be present to interpret the

report as necessary, as well as to help the committee focus on the judgements that are necessary.

Such a procedure smacks of that well worn Canadian custom, the Royal Commission, except that its efficiency would be greatly increased by having all of the data at hand in predigested form ready for the committee's action. Final recommendations would be forwarded to the Minister of Education for action.

The purpose of the example has been to show how some of the methodology of evaluation can be applied to specific circumstances. The aim of evaluation is to put curricular decisions on a rational footing. It is acknowledged that since evaluation is intrinsically associated with values, judgements will be necessary. The attempt is to make the rationale for these judgements public. Of course there are many problems that have been glossed over in the example, not the least of which is the lack of reliable measuring instruments. Nevertheless by systematizing and improving our evaluation procedures, we run less risk of making curricular decisions that are harmful to the students.

References

Anderson, G. J. The assessment of learning environments: A manual for the Learning Environment Inventory and the My Class Inventory. Atlantic Institute of Education, Halifax: 1971.

Caldwell, B. M. Preschool inventory. Princeton: Educational Testing Service, 1967.

Carkhuff, R. R. Helping and human relations: A primer for lay and professional helpers. New York: Holt, Rinehart and Winston, 1969.

Engelmann, S. The basic concept inventory, Pupil's test and scoring booklet. Chicago: Follett, 1967.

Frostig, M. Developmental test of visual perception. Palo Alto: Consulting Psychologists Press, 1963.

Kamii, C. K. Evaluation of learning in preschool education Chaper 13 in Handbook on Formative and Summative Evaluation of Student Learning. B. C. Bloom, J. T. Hastings and G. F. Madeus. New York: McGraw-Hill, 1971.

Moss, M. H. Test of basic information Unpublished text. George Washington University, 1967.

Mosychuk, H. Home environments and mental ability patterns. Unpublished doctoral dissertation. University of Alberta, 1969.

Provus, M. M., Kresh, E. and Green, B. F. Manual for administering the Preschool Academic Skills Test. Unpublished text. Pittsburgh Public Schools, 1968.

Roach, E. G. and Kephart, N. C. The Purdue Perceptual-Motor Survey. Columbus: Merrill, 1966.

Stake, R. E. The countenance of educational evaluation. Teachers College Record. 1967, 68, 523-540

Vinter, R. D.,Sarri, R. C., Vorwaller, D. J., and Schafer, W. E. Pupil Behavior Inventory. Ann Arbor: Campus Publishers, 1966.

CHAPTER IV

A Systems Approach

Introduction

One interesting and not unexpected development in the field of evaluation has been an emphasis upon the systems approach to evaluation of programs. The UCLA group, Stufflebeam and his co-workers as Ohio State and others have made extensive use of systems thinking in their work. What seems characteristic of this school of thought is an attempt to include more traditional notions about evaluation within a framework that comprehends a large number of variables or factors which have an impact upon curriculum programs.

In its simplest terms, this approach uses the concepts of input, output, and process or throughput to help define the domain of evaluation. In addition, and it is probably in this sense that this approach is an "administrative" or managerial approach, there is an emphasis upon an implementation of program phase and, over-all, a clearly defined decision situation for which evaluative information is required. Examined within the historical context of developments in management science, the approaches to evaluation referred to here are really particular applications of systems analysis to curriculum development and implementation problems. A thorough foundation in systems thinking would be useful, indeed almost mandatory, for educators and others who would apply these approaches to particular curriculum programs. Fluency in the language of systems analysis and the ability to use its various elements is therefore assumed as a pre-requisite to the motions discussed in the remainder of this section of the study. In the list of references, Churchman's book on the systems

approach (1968) is suggested a basic primer of ideas on this topic.

The Provus Approach. Perhaps the most useful example of the systems approach to evaluation is to be found in the work of Malcolm Provus. His work is eclectic in that it relies heavily on the various individuals who have done the pioneering work. However, Provus' framework for analysis of the evaluation process seems particularly strong in that it accounts for the concern explicit in the work of Robert Stake and the other so-called "non-administrative" workers in the area. Moreover, his analysis seems readily applicable to real life problems in curriculum development and implementation as they have appeared and are likely to appear in Alberta during the foreseeable future.

For Provus, an evaluation cycle is composed of four distinct stages: (a) definition, (b) installation, (c) process, and (d) product. He suggests that a cost-benefit analysis can be used as a final or supplemental stage in the cycle. The model which he proposes is all based on the concept of "discrepancy"; that is a difference between an established standard of performance and the actual performance of a program at any stage of its development and implementation.

Again, the concept of performance standard is obviously drawn from systems analysis. Compatible with the general notions of system analysis is the fact the performance standards are not irrevocably fixed *a priori*, but are subject to redefinition and modification in the light of experience with operation of the program or instructional system. The flow chart depicted in Figure 1 illustrates the Provus discrepancy model.

When discrepancy information is obtained, at least four distinct

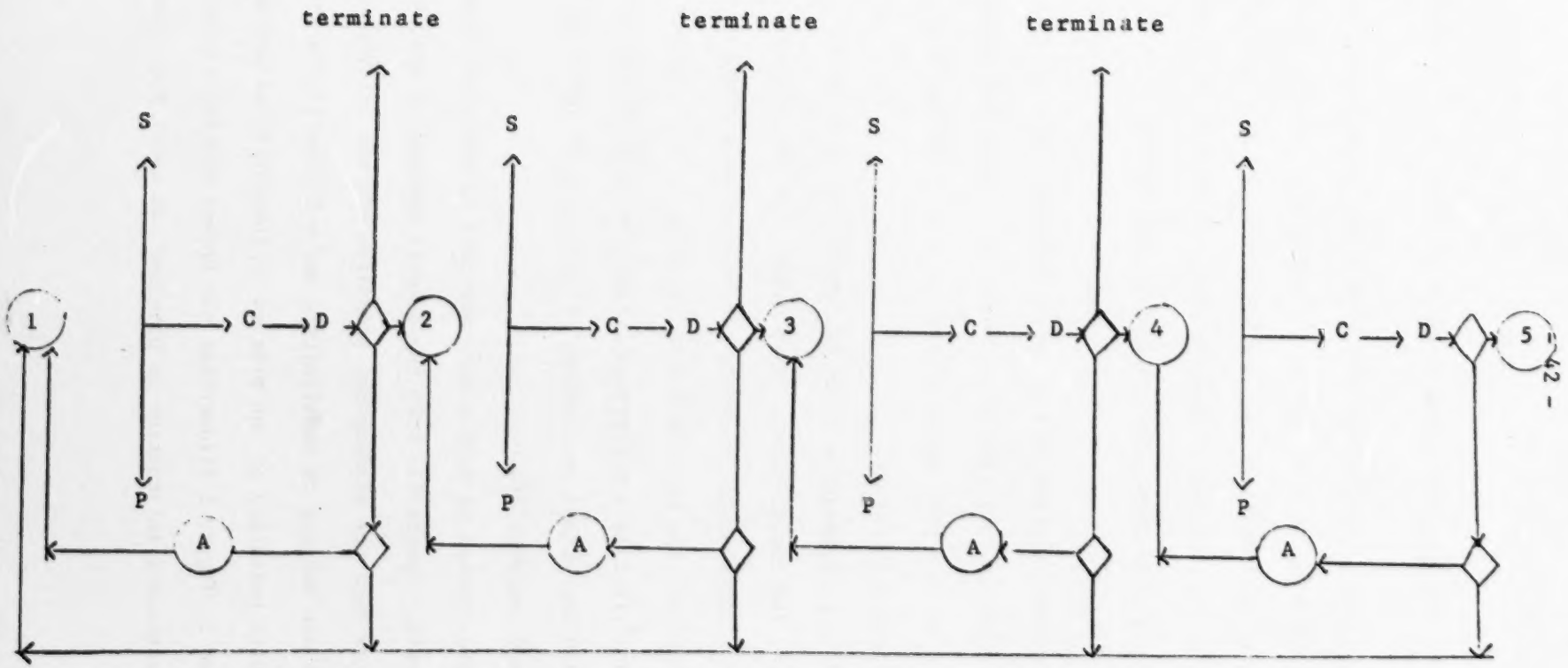

Fig. 1 - A flow chart designed to facilitate comparisons of program performance with standards.

S = standard, P = program performance, C = compute, D = discrepancy information, and A = change in program performance or standards.

decisions are possible: (a) go onto the next stage, (b) recycle the stage after changing the program standards or operations, (c) recycle to the first stage, or (d) terminate the program. As far as the four stages are concerned, they subsume some or all of the stages in the Stake model (antecedents, transactions, consequences), the Stufflebeam CIPP model (context, input, process, product), and also take account of the emphasis upon "installing the independent variable" that is found in much of the literature.

The Four-Stage Cycle.

Definition of the program content is the first stage of development. Based on a program-content taxonomy (such as the one developed by Stake, 1967) a definition of the particular instructional program is developed. Comparison between the defined program and the taxonomy will reveal information leading to one or another of the decision options referred to above. If, for example, the program definition takes inadequate account of the nature of inputs in terms of teaching staff qualifications, stage one will need to be recycled in order to obtain a revised program definition which does specify this particular input. In Figure 2, the taxonomy of program content is displayed. In a given situation, the criteria for adequacy of the program definition may or may not be as elaborate as those suggested by Stake. In any case, inclusion of this stage in an evaluation model emphasizes the important role that evaluators should play at the very beginning of program planning. Acceptance of the Provus model does not preclude evaluators from becoming involved after the program has been defined; but it does suggest a change from current practice whereby evaluation is not accounted for at early stages of program planning sequences.

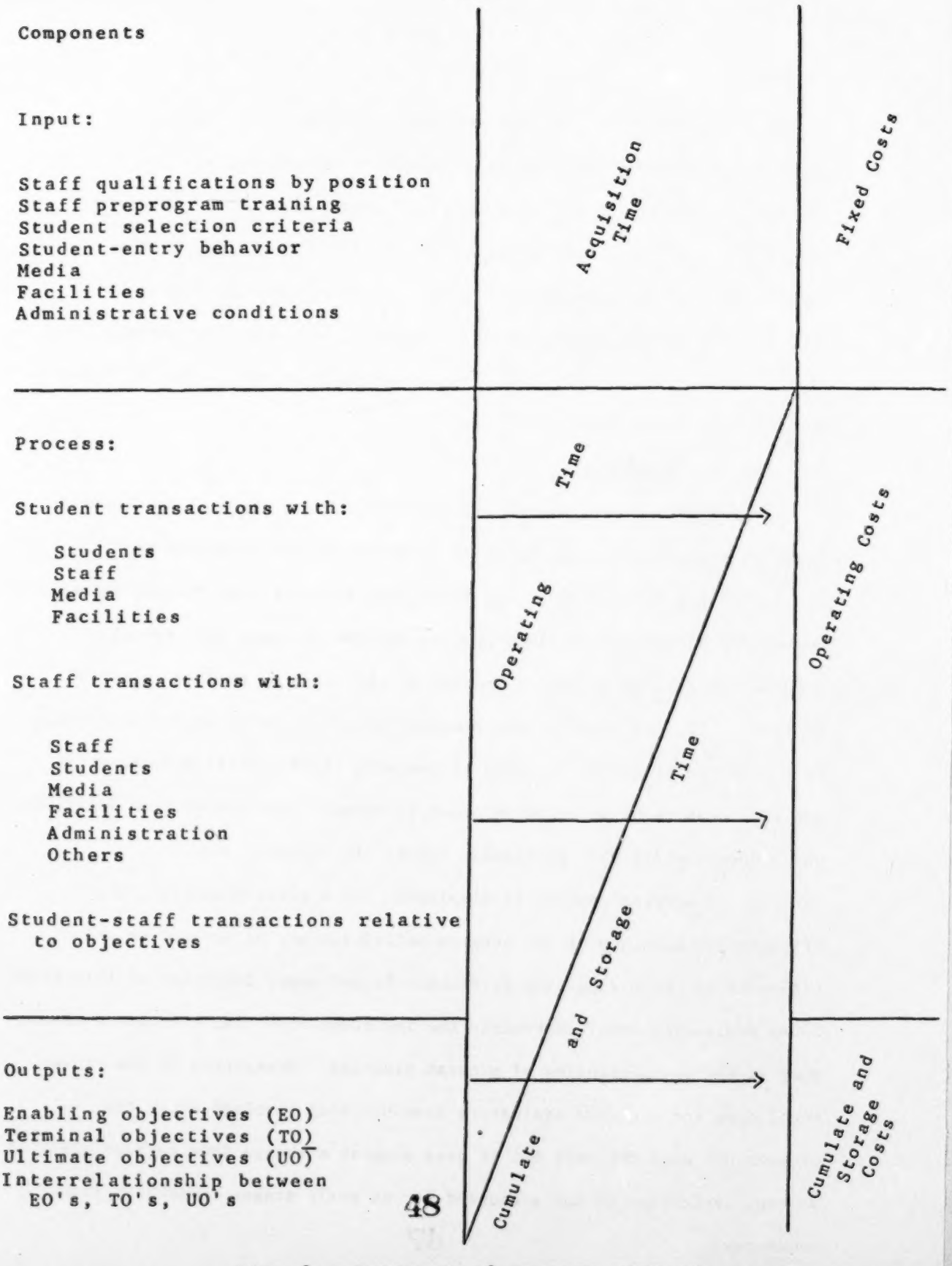

Fig. 2 - Taxonomy of program content

In the Installation stage, the performance standard is the program definition evolved during the first stage. Observations of actual program performance are compared with the program as defined and the discrepancy information resulting from the comparison is used for making a decision. For example, if the program calls for assignment of teachers to pupils on the basis of congruence in cognitive style and if no account has been taken of this factor in the on-going program, a re-cycling is required. The decision options would probably be either to recycle stage two and, after using tests of cognitive style, reassigning teachers and pupils, or to reexamine the feasibility of maintaining this aspect of the original program and possibly redefining the program so as to leave out this set of values. In other words, during these early stages of the evaluation and program cycle, modification in standards is always a possibility as empirical evidence establishes flaws in the original conceptualization of the program. Clearly, Provus is dealing with formative evaluation in this part of his model.

Stage three is process evaluation wherein discrepancies between what goes on in the program and the "enabling objectives" are obtained. Enabling objectives, following Stake, are interim or short-term or immediate indicators of the effect that the process of instruction has had upon pupils, teachers, groups, et cetera. If one takes the case of an open-space school, one result of the process of using large and small group organization as well as the traditional 25 - 30 pupil class of pupils may be that cliques are formed which remain intact even during large group activities. If this result is seen as contributing to or "enabling" the attainment of ultimate goals of the program, then this observation suggests no discrepancy on this

point and movement to the next stage is indicated. If, however, "a highly cohesive large work group" was, for whatever reason, seen as a desirable outcome of the process of instruction, the discrepancy information will lead to one or another of the other available decisions.

Finally, in stage four, product evaluation compares criterion measures applied to parts of the actual program with terminal objectives specified in the original program definition. If the definition specified a certain minimum level of performance on a standardized test by all pupils enroled in the program, determination of discrepancies will lead to one or another of the available decision choices. Recycling of any or all of the previous four stages is possible, if the results are extremely "poor" (i.e. if discrepancies are large), the program may be terminated. On the other hand, a redefinition of the program in terms of the enabling objectives may be proposed as a solution. Simply recycling some pupils through stages three and four may be all that is required in some cases; in the typical programmed learning sequence the application of this notion is well exemplified.

The inclusion of cost-benefit analysis as a possible additional stage will depend on the availability of alternative instructional programs each with the same or similar sets of objectives. If, for example, two or more programs are evaluated under each of the four stages described above, comparisons can be made on a stage by stage basis between or among the programs in terms of costs and benefits. Using an efficiency criterion, for example, one program may be superior to another not because achievement of terminal objectives is different; but because the installation stage is less costly in terms of recycling or "debugging" expenses. In terms of quality, achievement of enabling and/or terminal objectives may distinguish between two programs of equal cost.

Conclusion

The systems model described above seems worthy of examination by Alberta educators. Its utility in this setting cannot be known until it has been used in a variety of contexts. The logic of the approach is compelling. However, the paucity of educators, legislators, and members of the general public who are prepared to think in systems terms makes implementation of this approach a very difficult task. Some of the suggestions made in the final part of this study are intended to deal, to some extent, with this aspect of the problem. These suggestions tend to be fairly long-term in nature, dependent as they are upon training of evaluators and users of evaluation. On a shorter-term basis, agencies which are likely to become involved with changes in education in this province can begin to implement this systems model on a limited basis.even during the next few years.

An Example

As indicated in our Progress Report of February 1, 1971, we thought it useful to describe one or two actual exercises in evaluation with which we or our colleagues in the University of Alberta have been involved. One such description was included in the previous section of this study. Another example of evaluation methodology applied to an Alberta program is contained in the Hersom and MacKay (1971) study of open-area schools in the Edmonton Public School District. Because this latter study was sponsored by the school district it is not possible to disclose details of the findings or recommendations at the time of preparing this study for the Planning Mission. However, some generalizations which seem pertinent to some of the points already made maybe of use. It should be noted that these generalizations

are partly impressionistic in nature; but the consensus among those involved with the exercise is, to some extent, represented by the following statements.

1. Curriculum projects are often organized without due regard for evaluation at the various stages of development specified by Provus and others.
2. There continues to be a failure to distinguish between formative and summative roles of evaluation.
3. The thinking of both professional and lay persons associated with school systems continues to be affected by exposure to the traditional educational research design with its emphasis upon scientific generalizations. Evaluation that does not provide final or summative statements about success or failure is not widely accepted.
4. The installation stage, in Provus' terms, seems to be crucial in curricular innovations such as "open-space" schools. Beneath the label are many variants in instructional organization, pupil grouping, learning activities and so on.
5. Cost-benefit concepts are capturing the attention of school people. However, the other stages in an evaluation cycle are being neglected. The result is an overly simple emphasis upon input-output relations that are almost irrelevant to the process and goals of a particular program or project.
6. Many of the difficulties associated with a particular program may be attributable to managerial failure rather than to the internal functioning of the program itself.

References

Churchman, C. W. The Systems Approach New York: Delacorte Press, 1968.

Cronbach, L. J. Course improvement through evaluation. Teachers College Record 1963, 64, 672-683.

Hersom, N. L. and MacKay, D. A. A Study of Open-Area Schools in the Edmonton Public School District. Edmonton: Program Surveys Associates, 1971.

Provus, M. Evaluation of Ongoing Programs in the Public School System. Educational Evaluation: New Roles, New Means, in Tyler, R. W. (ed) Sixty-eighth Yearbook of the National Society for the Study of Education Part II, University of Chicago Press, 1969

Stake, R. E. The countenance of educational evaluation. Teachers College Record, 1967, 68, 523-540

Stufflebeam, D. L. Toward a science of educational evaluation. Educational Technology. 8, 5-12

CHAPTER V

Projection of Needs

The state of theory and methodology of evaluation is sufficiently advanced to support the suggestion that educational planners in Alberta would be well advised to place special emphasis upon the development and utilization of evaluation capabilities in this Province. The process of educational planning itself has imbedded in it an emphasis upon systems assessment and other of the components of evaluation. Certainly if change and innovation are to occur, the application of evaluation techniques at all stages of program development and implementation will be important for an educational system which is orienting itself to the future. Even if significant change were not to occur, the maintenance of effective educational programs and institutions is dependent in part upon the quality of information generated for decision-makers.

Our position is that if Alberta's present and, even more strongly, future needs in education are to be met, a number of goals should be accepted. These are:

1. That all agencies engaged in educational planning install evaluation as a sub-function of their operation.
2. That a center for the study of evaluation, modelled on those developed on the American scene, be established at an Alberta university, and that the mission of this center would include:
   (a) The development and adaptation of models of evaluation appropriate to the Alberta situation.
   (b) The provision of consultative and field services for all educational institutions in the Province.

(c) The preparation, through graduate degree programs, of persons qualified to work as evaluation specialists.

(d) The reeducation, through in-service education programs, short courses, and the like,of practitioners in various kinds of educational organizations.

(e) The dissemination of information and points of view on evaluation to members of the public, to government, and various sectors of the population of Alberta.

Accordingly, the provincial government should invite groups from the Universities to submit proposals for establishment of such a center and provide the funds necessary to ensure its successful operation. Even if conceived on a very modest scale, the contribution of such a group could, in our opinion, make a significant contribution to the improvement of education in Alberta.

APPENDIX A

ALBERTA DEPARTMENT OF EDUCATION

REQUEST FOR PROPOSALS

Nos. 1970 - E - 2
1970 - E - 3

CONCERNING

A PRESCHOOL EDUCATION PILOT PROJECT

IN EDMONTON AND CALGARY

ISSUED BY

ROBERT C. CLARK

MINISTER OF EDUCATION

OBJECTIVES

The Government of Alberta desires to contract with responsible persons and organizations to achieve the following objectives:

1. To select a representative group of disadvantaged children from the inner-city core of the City of Edmonton and the City of Calgary. These children should be eligible to enter Grade One the year following their acceptance into the program.

2. To identify the nature of the handicaps of each child.

3. To design an appropriate program of personal development for these children. The aim of this program will be to enable each child to adapt successfully to the demands and opportunities of elementary school life.

4. To carry out this program over a two year period.

GENERAL GUIDELINES

Responses to this RFP, and any subsequent project carried out by a contractor, must be governed by the following guidelines:

1. Children to be served by the project must be culturally handicapped but physically normal. "Culturally handicapped children" may include:

   a) Children from very low income homes.
   b) Children from broken homes.
   c) Children who receive little love or attention.
   d) Children whose parents speak a minority language only.
   e) Children who lack experience working and playing with others.

2. Children selected for the project should form a heterogeneous group with respect to sex, race, religion, and background of parents.

3. Children selected must be children who likely will enrol in the Alberta educational system in the following year.

4. Proposals (and projects based upon them) must make

specific provision for an initial medical examination of all children selected.

5. Proposals must describe how parents will be consulted and involved in the design and implementation of the project.

6. Proposals must describe the curriculum to be followed.

7. The contractor will be allowed complete freedom with respect to the hiring of staff.

8. The contractor will be required to maintain adequate insurance. Questions regarding insurance should be directed to: Director of School Administration
Department of Education
628 Administration Building
10820 - 98 Avenue
Edmonton 6, Alberta.

9. The contractor will be required to keep complete and accurate financial and service records, and to make these available for inspection upon request.

10. Facilities used must conform with the standards laid down in the Welfare Homes Act. Meals must conform with regulations outlined in Standards for Institutions and Nurseries, Department of Public Health, Government of Alberta. Regulations under the Fire Prevention Act, Institutions and Nurseries must also be complied with.

11. Proposals will be evaluated on the basis of:

    a) Creativeness and practicality.

    b) Cost effectiveness (how much implementation of a given proposal will contribute toward the achievement of the objectives per dollar spent by the Government of Alberta).

    c) Second order costs and benefits.

    d) Conformity to guidelines.

12. Project evaluation will be conducted on the same basis.

13. No proposal submitted in response to this RFP is necessarily accepted. In the event that no single

proposal is acceptable, the Government of Alberta reserves the right to synthesize the best features of all responses and to re-tender the project.

14. A two year contract will be offered to the successful respondent, with the possibility of a one year renewal at the end of that period. The maximum time period allowed for this pilot project is three years.

15. Contracts may be cancelled for non-achievement of objectives, or for failure to adhere to these guidelines.

16. Inquiries concerning any aspects of this RFP are welcome, and should be directed to:

Dr. E.J.M. Church
Director of Special Educational Services
702 Administration Building
10820 - 98 Avenue
Edmonton 6, Alberta.

FINANCIAL GUIDELINES

1. No financial support will be provided for the purchase or construction of buildings or physical space. The Government is anxious to see the largest possible proportion of funds flow into operating costs and instructional supplies.

2. The Government of Alberta is prepared to pay up to $50,000 per year toward the attainment of the objectives of this RFP.

3. Organizations submitting proposals may be able to arrange additional financing or contributions-in-kind from other sources. While not essential, such additional financing and contributions will be considered favorably in the evaluation of responses to this RFP.

TIME SCHEDULE

1. Responses to this RFP should be submitted to:

Dr. E.J.M. Church
Director of Special Educational Services
Department of Education
702 Administration Building
10820 - 98 Avenue
Edmonton 6, Alberta

by May 25, 1970.

2. It is anticipated that a contract should be negotiated with the successful respondent by July 2, 1970.

3. The project should commence on or about September 1, 1970.